INDIVIDUAL PERSPECTIVES TOWARDS DIMENSIONS OF COLLEGE TEACHING EFFECTIVENESS

A THESIS

Presented to

The Faculty of the Division

of Graduate Studies

By

Leon James Tauscher

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Psychology

Georgia Institute of Technology

May, 1975

INDIVIDUAL PERSPECTIVES TOWARDS DIMENSIONS

OF COLLEGE TEACHING EFFECTIVENESS

Approved:

Sam C. Webb, Chairman

Gerald W. Day

Lester A. Neidell

William W. Ronan

Date approved by Chairman: May 27, 1975

ACKNOWLEDGMENTS

The writer wishes to thank the members of his Thesis Advisory Committee, Drs. G. J. Day, L. A. Neidell, and W. W. Ronan, for their timely assistance and guidance in conducting this study. Sincere gratitude goes to Dr. Sam C. Webb, chairman of the committee, for his patient and skilled efforts in directing the research, editing the many written products, and especially for helping to inspire in me a sincere attitude to the conduct of psychological research. Special thanks are also expressed to the faculty and graduate students of the School of Psychology, and especially to the ninety-four individuals from all levels at Georgia Institute of Technology who graciously gave of their time to participate willingly in this research. To this writer's personal friends, Michael H. Frisch and Garland L. Wright, goes special gratitude for their suggestions, clerical assistance, and moral support rendered to me throughout the duration of this research. In this same respect, thanks go to Joan M. Swetz, Agnes V. Stelzner, and Morrill E. Stelzner, who each in their own way contributed many hours of assistance in completing this study.

This thesis is dedicated to my wife, Joan, and to our children, Brian and Angela. Of all who contributed to this research, none gave so graciously of themselves as did the members of my family.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

SUMMARY

Because of the theoretical and practical importance of performance ratings, it is necessary to determine why different raters observing the same performance typically do not agree on their evaluations of that performance. Carroll & Chang's (1970) individual differences multidimensional (INDSCAL) model was used to evaluate the hypothesis that individuals from different levels within the same academic organization do not attribute importance to dimensions of college teaching effectiveness in the same manner.

The most significant finding was that eight distinct clusters of individuals were identified on the basis of the differential importance each attributed to behaviors comprising perceived dimensions of teaching effectiveness. Combining subjects on the basis of actual academic positions resulted in a marked reduction in the magnitude of group differences in performance perspectives, and changed the nature of the research results and interpretations that could be ascribed to them.

The general conclusion from this research is that individuals differ considerably in the manner in which they attribute importance to dimensions of college teaching effectiveness. The overall findings offer general but indirect support for Guion's (1965) contentions regarding

the causal bases underlying disparities of performance ratings.

CHAPTER I

INTRODUCTION

Many areas of psychology are heavily dependent on subjective measures of behavior as the basic data of research and application. In particular, ratings are the most frequently used means of evaluating performance (McCormick & Tiffin, 1974). They have been described by some writers as being valuable, expedient, comprehensive, and in some cases the only behavioral measures available (Anastasi, 1950, 1968; Campbell, Dunnette, Lawler, & Weick, 1970; Smith & Kendall, 1963). Yet, they have also been shown repeatedly to lack reliability and validity, especially when the index of inter-rater agreement is used to infer reliability and/or validity (see reviews by Odiorne, 1964; Ronan & Prien, 1966, 1971, 1974; Ronan & Schwartz, 1974).

Assessment of inter-rater agreement always involves two or more raters or rater groups evaluating the same performance. Indices of such agreement are obtained through processes comparable to parallel forms assessments of test reliability. A general pattern that emerges from studies assessing agreement between raters within the same group and between two or more rating groups is that between group indices are considerably lower than those obtained between

members of the same group. A study by Springer (1953) provides a good example. She obtained within group reliabilities ranging from .34-.48 for peers and from .56-.71 for superiors who rated foremen on eight performance dimensions. Between group correlations for the same two rating groups, however, ranged only from .15-.39 for the eight dimensions. While the levels of both types of inter-rater indices cited are frequently found in more recent studies, it is not uncommon to find studies reporting even lower, zero, and negative correlations. Basco & Lawshe (1959), for example, obtained correlations ranging from -.15 to .08 between superior and subordinate ratings of leadership characteristics in foremen; and Ronan & Latham (1974) reported correlations ranging from -.52 to .20 between discriminant functions based on ratings of forresters and dealers regarding 14 dimensions of pulpwood producer performance. These latter results are surprisingly low since both rater groups worked closely with the performers being rated, had frequent opportunity to observe their performance behaviors, and provided ratings based on actual observed frequency of specific behaviors.

Such low relationships between ratings of groups of raters who are considered equally appropriate and qualified for assessing the performances of a set of people seriously impair the credibility and utility of such ratings in applied research. Further, they raise serious questions concerning

theoretical aspects of behavior observation. As a first step towards improving both types of inter-rater agreement, it is essential to discover why different individuals observing the relatively same performance of a given person cannot agree on what they are observing.

Ronan & Prien (1974) suggest two major sources of variance associated with the assessment of performance: (1) true performance variability specific to the performer, and (2) observational variability specific to the observer and the measurement of the performance. While the intent of applied research is to assess true performance variability, Ryans & Frederickson (1951) have pointed out that the accuracy of such assessment is necessarily dependent upon the reliability and validity of the observation and measurement of the performance. When ratings are used to measure performance, the rater himself becomes the sole determinant of the accuracy with which the true performance is assessed.

In making a rating, the rater is required to observe the behavior of another person and then make a judgmental response. However, most research on ratings has given little attention to the rater as a source of variance (Blum & Naylor, 1968; Guion, 1965; Ronan & Schwartz, 1974). Rather than being concerned with the behavior of the rater <u>per se</u>, research on ratings has been almost solely concentrated on factors and variables external to him. Some investigators, for example, have focused on assessment methodologies as a

source of error (Smith & Kendall, 1963); and others (Ghiselli, 1956; Schmidt & Kaplan, 1971; Wherry, 1957) have consistently emphasized the lack of proper criterion conceptualization.

That these factors do not appreciably account for the low reliability of ratings can be illustrated by research on student ratings of college teachers. Many factor analytic studies performed with a variety of methodologies and diverse student groups have produced first order and second order factors that are remarkably similar (Coffman, 1954; Deshpande, Webb, & Marks, 1970; Solomon, 1966; Solomon, Rosenberg, & Bezdek, 1964; Thompson, 1972). But despite this knowledge of the dimensionality of college teaching effectiveness, studies on rated teacher performance, like industrial studies, have resulted in low, zero, or negative inter-rater reliabilities (Borg & Hamilton, 1956; Braunstein & Benston, 1973; Buckner, 1959). For example, Braunstein and Benston reported almost total disagreement between student ratings and department head rankings of faculty members on five criteria. Of 27 of the rank order correlations computed, 16 were negative. While the remainder ranged from zero to .60, the approximate median was close to zero. Similarly, Rodin & Rodin (1972) obtained a correlation of -.754 between student ratings of teaching assistants and the objective criterion of actual student learning of calculus.

Many years of research have not provided an understanding of why these types of results continue to be

obtained. However, one potential source of error not yet investigated successfully has been the nature of the rater himself. This study attempts to investigate one rater characteristic, namely the valued importance of job tasks, as a possible source of differences among performance ratings.

CHAPTER II

REVIEW OF THE RELEVANT LITERATURE

It has been noted above that between group reliabilities tend to be lower than those between individuals within the same group. This pattern suggests that different individuals belonging to different rater groups somehow differentially evaluate the same task performance. In treating the general topic of inter-rater reliability, conventional psychometric assessment assumes that the target performance is evaluated in relatively the same manner by all raters (Ronan & Prien, 1974). If different raters do not evaluate the same performance in the same manner, then it would be understandable that their ratings would not agree well, if at all.

All investigators of inter-rater reliability seem to make either explicitly or implicitly, the above assumption in conducting their studies. Generally in performance evaluation, different raters are given identical assessment instruments that contain behavioral stimuli that can range from global dimensions, such as competence, to specific behaviors, such as comes to work on time. The raters are then assumed to evaluate the different performers on the specified stimuli using the same relative underlying

behaviors for all performers in the same manner. Because this requisite condition is assumed rather than applied, it may be appropriately called the "identity of evaluation" assumption associated with the assessment of inter-rater reliability. Yet, the data already cited would seem to suggest this is not a valid assumption.

Variables Affecting Subjective Performance Evaluation

There is in fact much evidence to suggest that a host of personal and situational variables differentially affect the manner in which individuals observe and evaluate the performance of another. Most salient among these have been personality characteristics (Graham & Calendo, 1969), rater attitudes towards behavioral constructs used on the rating instrument (Eiser, 1973) and towards the performer (Thorndike, 1949), cue saliency associated with the attribution of causality to behavior (Beckman, 1970), and the personal relevance of the rating to the rater (Callahan & Messé, 1973; Ronan, 1970). In general the studies noted that ratings were systematically distorted or biased in direct relation to the respective variables. That is, in each study a significant portion of the total rating variance was accounted for by one or more factors extraneous to the true performance being assessed. For example, Thorndike (1949) found that while instructor pilots rated the trait of "likableness" as least important to success in flight school, the trait

correlated highest of all with the instructors' ratings of general flying ability of their cadets. Even though extraneous to true performance, likableness nonetheless accounted for more variance than any of the other traits being rated.

Rater perspectives represent a general category of variables that also have a significant effect on the ratings people make. As a basic minimum, the identity of evaluation assumption would require that different raters generally have common perspectives regarding: (1) what the target performance is, (2) what behaviors, traits, or characteristics determine performance effectiveness, and (3) what the relative standards of effectiveness are for the given job. Available evidence, however, suggests there are extensive differences among individuals regarding their perspectives towards each of these three areas.

A general finding has been that the nominal assignment of an identical title to a target performance does not make it mean the same thing to different individuals sharing the same job context. Stogdill, Shartle, Wherry, & Jaynes (1955) found that executive Navy officers accomplished general administrative duties in many different manners. Depending on their designated administrative functions, the officers differed in respect to what behaviors they performed and in the relative importance such behaviors played in determining job success. Similarly, Roach & Wherry (1970) found that

insurance salesmen could be differentially identified on the basis of their perceptions regarding the importance of six types of selling behaviors associated with six different types of insurance policies being sold. These findings suggest that asking a rater to judge someone's "administrative" or "salesman" effectiveness might elicit ratings based on concepts of the nature of the job or behaviors being rated that differ from rater to rater.

If different raters have different concepts regarding the target performance, then it is also likely they will differ in their perspectives regarding the determinants of performance effectiveness. Individuals within the same job context have been found to disagree in their perceptions of what the specific duties of a job are (Prien, 1962), what specific behaviors are critical to job success (Ronan & Latham, 1974), and what characteristics are most important for job success (Maslow & Zimmerman, 1956; Stander, 1965). Other researchers have reported basic disagreements regarding the perceived importance of personal traits for managerial success (Gruenfeld & Weissenberg, 1974) and the related importance of organizational variables in determining actual performance (Friedlander, 1966) and in determining job satisfaction (Schwartz, Ronan, & Day, 1975). A finding that differs somewhat from these was reported by Crawford & Bradshaw (1968). They found that students, teachers, and administrators agreed quite well on the ranked importance of

teacher behaviors to effectiveness. However, the agreement was only relative, as a subsequent analysis revealed that there were marked differences in the magnitudes of importance assigned to the behaviors by the different groups.

Further evidence that the perceived underlying determinants for success are different for various individuals or groups comes from factor analytic studies using different rater groups and variables to evaluate the same performance (Klimorski & London, 1974; McClelland & Rhodes, 1969; Ronan, 1963a, 1963b; Turner, 1960). Each of these studies resulted in the rater groups being identified as orthogonal factors. Moreover, the factor loadings of the rated variables differed markedly for the different groups of raters, despite the fact they were evaluating the same persons on the same job. What these findings suggest is that perspectives towards the dimensionality of a given job, and the perceived determinants of that dimensionality, are quite different for different individuals or groups.

The evidence, then, strongly suggests there <u>are</u> basic differences in the manner in which individuals or groups approach and evaluate the same job performance. Many investigators have discussed this problem in general and have offered arguments as to why such notable differences exist.

Plausible Bases for Differences in Performance Perspectives

Schrauger & Altrocchi (1964) have noted that the

behavior of one individual serves as informational cues to others so that each cue seemingly carries differential amounts of information and value for different observers. While Ronan & Prien (1974) have provided extensive detail on the differential effects that personal, contextual, and situational variables have on different workers within the same job environment, others (Barrett, 1966; Guion, 1965) have suggested that a person's position in an organizational hierarchy may significantly determine his perspectives towards the performance dimensions of a given job. Individuals at different levels serve different functions, and have different responsibilities, wants, needs, and values. Guion (1965) specifically suggests that because of these differences, raters in different organizational positions attend to differential job behaviors, differentially relate and value these to different job performance dimensions, and differentially relate and value these dimensions to global job effectiveness. Further, he suggests these differences might serve as the causal basis for noted disparities of ratings.

While this problem has been discussed extensively, research relating to the potential bases for differences in perspectives is practically non-existent. As previously cited, Crawford & Bradshaw (1968) found that individuals from different levels within a university showed marked differences in the magnitudes of importance assigned to teacher behaviors

related to effectiveness. Similarly, a recent study by Zedeck, Imparato, Krausz, & Oleno (1974) considered the extent to which nurses at different organizational levels differed in the values they attributed to specific performance behaviors. Head nurses and registered nurses (both n's = 6) rated each of 249 reported critical nursing behaviors on a seven point scale of importance. Of the 249 behaviors, 177 of them (77%) were assigned higher mean values by the subordinate nurses (sign test, $p < .001$). Although the rating groups were too small to determine quantitative value differences, the suggestion is clearly present that the two groups were not assigning the same value to the same performance behaviors.

These studies tend to offer general support for Guion's (1965) arguments regarding basic rater differences in perspectives towards specific performance behaviors. However, the research to date offers little insight as to whether or not similar differences exist regarding the manner in which individuals at various organizational levels view the importance of different job dimensions towards determining performance effectiveness. If such differences were shown to exist, then further support would be provided for Guion's arguments regarding the differential causal bases underlying the disparities of ratings.

CHAPTER III

THE THESIS

The research evidence cited concerning the low level of inter-rater agreement, especially that between groups of raters, indicates that the relatively same performance is somehow differentially evaluated by different individuals or groups of seemingly equally qualified raters. Past findings also indicate that a major source of variance in such ratings is associated with differences among raters in respect to a variety of personal characteristics. One such characteristic that has been recently noted relates to the possible differences among raters at different organizational levels in respect to the extent they may value differently the behaviors being rated. Much more thorough and in depth research is needed, however, before individual value systems can be established as an underlying basis of differing rating perspectives. This study seeks to contribute to this needed research by examining in an exploratory way the extent to which people from different organizational levels differ in their individual systems of valuing the importance of relevant job behaviors.

Since most jobs are multidimensional in nature, the study considers a behavior domain that has been clearly

identified as multidimensional, namely college teaching effectiveness. Thus in assessing the extent to which individuals differed in their systems of valuing the importance of relevant performance behaviors, the following hypothesis was evaluated:

> Individuals from different levels within the same academic organization do not attribute importance to dimensions of college teaching effectiveness in the same manner.

To examine the hypothesis, individuals from different levels within the same academic organization provided a complete rank order of 30 relevant teaching behaviors with respect to their relative importance in determining overall college teaching effectiveness. These data were then analyzed using Carroll & Chang's (1970) individual differences multidimensional scaling (INDSCAL) model, which offered the unique advantage of looking at clusters of teaching behaviors (i.e., dimensions) and individual locations, simultaneously, in the same multidimensional space. The importance of this common reference space containing both teaching dimensions and subjects was that it made possible a direct estimation of the amount of importance the different individuals attributed to the derived dimensions of effectiveness underlying their actual judgments of specific behaviors.

CHAPTER IV

METHOD OF INVESTIGATION

Subjects

Ninety-four individuals from four levels within the academic structure of Georgia Institute of Technology volunteered to participate in the study. From this subject pool, a random sample of 10 subjects was selected from each of the following four organizational levels: (1) Administrators, defined as individuals occupying positions of School Director and Dean; (2) Teachers, defined as individuals occupying positions of Assistant Professor through Professor; (3) Graduate Students; and (4) Undergraduate Students. The sample was limited to 40 subjects because of computer programming restrictions, and consisted only of males.

Procedure

Data Collection

Approximately 400 specific teacher behaviors reported in previous research as being related to various dimensions of teaching effectiveness (Deshpande, et al., 1970; Isaacson, McKeachie, & Milholland, 1963; Ronan, 1971; Thompson, 1972) were compiled. From these the writer selected 63 which he considered non-overlapping in content and meaning and related to the following five commonly identified dimensions of

teaching performance (the number of items per factor are given in parentheses): (1) Inspires Creative and Critical Thinking (13); (2) Subject Mastery and Course Organization (15); (3) Rapport (13); (4) Overload (10); and (5) Evaluation (12). From the 63 behavioral items, 30 were selected for use in this study by randomly drawing them from a container. The 30 statements were numbered and typed on two sheets of paper. After reproduction, the items were cut apart and taped onto blank sheets of paper so they could be easily removed by the subjects during the experimental task.

The subjects were treated individually and allowed to work at their own pace. Each subject was given a three part booklet containing instructions and materials, along with a small cardboard box partitioned into five compartments. The compartments were indexed from left to right as Little or No, Some, Quite a Bit of, An Extreme Amount of, and An Extraordinary Amount of Importance, respectively. Using the box as an aid to sorting, the subjects rank ordered the 30 items from most to least important in determining overall college teaching effectiveness by use of a two step procedure (see Part 2 of Appendix A). Basically each subject read over all 30 of the statements, and then placed six items in each of the five compartments of the box according to their judged degree of importance. The subjects further rank ordered the six behaviors within each of the five categories from most to least important. The five sets of rankings were then

combined to produce the desired single rank order of the 30 behaviors.

The items were assigned integer values from one to 30, corresponding to their positions in the rank orders from least to most important. The rank order of items given by each of the 40 subjects is presented in Appendix B.

While not central to the thesis research *per se*, additional information was collected for possible use in studies that extend beyond the scope of this thesis (see Parts 1 and 3 of Appendix A).

Data Analysis

The rank order scores for the 30 teacher behaviors were transformed into Euclidean distances for each subject. Because only one rank order of the items was obtained from each subject, the distance between any two items *i* and *j* was equal to the square root of the squared difference between their respective rank order scores. This transformation, comparable to obtaining the absolute value of the difference between any two respective items, resulted in a 30 x 30 symmetric dissimilarities matrix for each of the 40 subjects.

The INDSCAL Model

These matrices were than analyzed by use of Carroll & Chang's (1970) INDSCAL model. A major feature of INDSCAL is that it allows one to assess directly the manner in which individuals differentially assign values to the major derived dimensions underlying their perceptions of a multidimensional

stimulus domain. More important, this individual determination is accomplished without imposing a conceptual structure on either the stimuli or the subjects. Thus the relationships between subjects and stimuli are analyzed simultaneously within the same multidimensional space so that whatever relationships emerge are a function of the perceptual behavior of the subjects rather than a function of the conceptual framework of the researcher.

The INDSCAL model provides three major types of output data. With respect to its application in this study, the first type is a matrix of item coefficients (analogous to factor loadings) which define the locations of items in respect to the dimensions of the evaluative space derived from the rank order scores. Second, it provides a matrix of subjective "weights" which define the locations of subjects in respect to these same dimensions. These weights also reflect the relative amount of importance the subjects attribute to the dimensions in the course of evaluating the items. Third, it provides two dimensional plots of: (1) the stimulus items with respect to their metric relations to the dimensions, and (2) the subjects with respect to their metric relations to the dimensions. All possible pairs of dimensions are plotted consecutively, using the coefficients from the two respective matrices noted above.

Another major characteristic of INDSCAL is that it was designed specifically for use in psychological research. It

has been used effectively in studies concerned with the assessment of individual differences related to conceptions of nations (Wish, 1971), subjective weights assigned to stimulus attributes of facial similarity (Tversky & Krantz, 1969), color perception (Carroll & Chang, 1970), and market research (Green & Carmone, 1970).

Hence the three assumptions of the model are closely aligned with empirical evidence regarding human perceptual and evaluative processes (see Carroll & Chang, 1970, pp. 314-318). The model assumes that the stimuli have a linear additive relationship to the derived dimensions of the space. That this assumption is highly robust with respect to multidimensional scaling techniques in general has been reported by Green & Rao (1972) and more recently by Green (1975). A second assumption is that for all individuals, r common dimensions underlie the n stimuli. This implies that the stimuli are representative of a common set of dimensions inherent in the behavioral domain. In this analysis, this assumption appears tenable since the stimulus behaviors were selected on the basis of their relationships to previously derived factors underlying college teaching effectiveness. The third assumption is that individuals differentially "weight" the common dimensions of a stimulus space. This assumption allows for but does not force individuals to differ with respect to the importance attributed to the common dimensions of the space. Theoretically

each subject could either "weight" all the dimensions the same, could attribute total importance to one dimension to the complete exclusion of the others, or could weight the respective dimensions in any other conceivable combination.

In studies using INDSCAL analysis, the general practice is to collect data by paired-comparisons techniques with the subjects being asked to indicate the extent to which pairs of stimuli are similar. The data collection procedure used here differed from the usual procedure in that the judgmental frame of reference was restricted to ostensibly one dimension (viz, importance) as the subjects gave only one ranking of the items. In some respects this method of data collection resembles that associated with the law of comparative judgment or the law of categorical judgment scaling models. Possible effects of this departure from normal procedure are noted later.

INDSCAL Analysis and Selection of a Dimensional Solution

The computer program for INDSCAL analysis used in this study was written by J. J. Chang and J. D. Carroll at Bell Telephone Laboratories, Murray Hill, New Jersey, in 1968, and adapted for the UNIVAC 1108 by L. A. Neidell, Georgia Institute of Technology, in 1969. With INDSCAL analysis, the user specifies the desired number of stimulus dimensions he wishes from each respective analytic solution and a criterion for stopping the iterative procedure used by the INDSCAL algorithm in attaining a simultaneous least-squares

solution for the parameters of the stimulus space and the subjects' locations in that space. When the error sum of squares associated with the parameters being estimated is no longer reduced by a value greater than the criterion value specified by the user, the iterative procedure is stopped.

Using the 40 dissimilarities matrices as input, the program was instructed to produce a three, four, five, and six dimensional solution of the stimulus space. The criterion value specified was .001, which was reached for all solutions in not greater than 40 iterations. Once the specified dimensional solutions were attained, the question arose as to which solution to evaluate.

A decision regarding which dimensional structure to select as most appropriate for this study was made on the basis of three criteria (personal communication with J. D. Carroll, 1975). The first criterion is associated with the "degree of fit" of the raw stimulus data to the derived stimulus space as reflected by the approximate percentage of variance accounted for by the different solutions. These data for each of the four different solutions are given in Table 1. As can be observed, the four dimensional (4-D) and 5-D solutions account for the largest percent of variance. The marked reduction in variance accounted for by the 6-D solution should not have occurred, and indicates that the INDSCAL algorithm converged to a local minimum. While this ruled out the use of the 6-D solution, the 4-D and 5-D

Table 1. Approximate Percentage of Total Variance Accounted for by Each Dimension within the Various Solutions

Solution	Dimension 1	2	3	4	5	6	Total
3-D	15.41	12.66	11.95				40.02
4-D	16.56	13.28	9.81	6.40			46.05
5-D	13.01	12.44	11.56	6.85	6.27		50.13
6-D	8.32	8.04	6.35	4.62	4.35	4.04	35.72

solutions were still possible alternatives.

The second criterion is the average correlation between computed scores and original data for the subjects. This criterion concerns the "degree of fit" of the subjects' actual locations in the stimulus space versus their predicted locations based on the subjective weights derived for them in a particular solution. Each subject's correlation, the average correlation, and the mean squared correlations for each of the solutions are given in Table 2. As can be seen, there are no marked differences in these data for the 4-D and 5-D solutions.

The last, most important criterion considers the interpretability of the dimensions of a solution within a meaningful psychological frame of reference. Unlike factor analysis, the axes associated with an INDSCAL solution cannot be rotated but must be interpreted directly. Like factor analysis, there is no absolute cut-off point between significant versus nonsignificant coefficients relating the stimuli to the derived dimensions. According to J. D. Carroll (personal communication), one should treat the stimulus coefficients as factor loadings (although in practice they are characteristically of lower magnitude) and interpret the dimensions like factor analytic dimensions.

Following this course, this writer attempted to interpret the dimensions within the 4-D and 5-D solutions. While both solutions produced meaningfully interpretable

Table 2. Correlation Between Computer Scores and Original Data for Subjects by Various Dimensional Solutions

Subject	3-D	4-D	5-D	6-D
1	.164	.659	.724	.714
2	.424	.457	.460	.683
3	.628	.598	.629	.607
4	.320	.669	.669	.678
5	.221	.196	.281	.549
6	.705	.702	.759	.761
7	.124	.650	.677	.661
8	.719	.709	.743	.819
9	.594	.579	.618	.600
10	.437	.840	.629	.834
11	.577	.592	.603	.608
12	.765	.783	.785	.857
13	.726	.750	.749	.801
14	.715	.725	.723	.763
15	.443	.428	.534	.538
16	.736	.772	.763	.762
17	.818	.856	.846	.845
18	.362	.408	.411	.487
19	.694	.719	.738	.745
20	.726	.733	.776	.769
21	.580	.540	.573	.612
22	.850	.854	.856	.868
23	.641	.636	.681	.701
24	.842	.872	.869	.913
25	.807	.857	.856	.862
26	.740	.790	.794	.803
27	.443	.432	.452	.487
28	.847	.867	.886	.876
29	.296	.328	.354	.358
30	.564	.571	.570	.603
31	.688	.632	.757	.808
32	.389	.380	.454	.472
33	.646	.624	.689	.671
34	.613	.619	.683	.729
35	.697	.696	.690	.696
36	.618	.737	.802	.801
37	.768	.769	.847	.853
38	.486	.628	.884	.888
39	.809	.784	.807	.842
40	.845	.850	.846	.848
AVE r	.602	.659	.691	.719
MEAN r^2	.400	.460	.501	.535

dimensions, the 4-D solution produced the more distinct and less ambiguously interpretable dimensions of teacher effectiveness. Thus the 4-D solution was selected for use in evaluating the hypothesis.

Treatment of the 4-D INDSCAL Data

As previously noted, the first treatment of the selected solution was the interpretation of the dimensions. While precise interpretation is not germane to this research, being able to distinctly identify and tentatively label the derived dimensions in terms of the teaching behaviors each encompasses is a helpful aid in understanding and interpreting the dimensions to which subjects attributed importance in ranking the behaviors.

The four dimensions were interpreted primarily in terms of the behaviors clustering at their polar ends, as these led to the most meaningful identifications. The complete set of teacher behaviors used as stimuli and their coefficients associated with each of the four dimensions are given in Appendix C. These coefficients will be referred to in the remainder of this analysis as "loadings." The absolute values of the loadings ranged from .006 to .360, and the minimum criterion selected for interpreting the dimensions was .200 or greater. Items with loadings lower than this did not enhance the interpretations.

In interpreting the dimensions, an absolute value (e.g., .350 or greater) was selected, and those items meeting

this value were extracted from the factor loading matrix (Appendix C) and listed under their respective dimensions. The absolute value was reduced by .01, and the procedure repeated iteratively until the minimum absolute value criterion of .200 or greater was reached (as shown in Table 3). At each step this writer interpreted each dimension with respect to the common attributes of the behaviors having loadings on the respective dimensions. In all four cases there emerged seemingly clear concepts characterizing or summarizing the behaviors so identified for the different dimensions. The titles assigned the dimensions were selected to describe these differential concepts.

After the dimensions were interpreted, the individual subjective "weights," representing each subject's relative location to the four dimensions, were analyzed. The six two-dimensional plots were first inspected to determine an overall picture of subject "scatter" about the dimensions. Then the plots were examined to detect if any subjects tended to cluster consistently about any one or more of the four dimensions. Subjects who formed such stable patterns were placed into groups, and average weights for each dimension for each group were computed and compared. In addition the average raw scores (ranks) for each of the 30 behaviors were computed for each identified group. These raw scores were then examined to determine whether the subjects within the clusters assigned similar raw score

values to the behaviors comprising the dimensions to which they respectively attributed most importance. As a last step, a comparison was made between the average group weights of the identified clusters of subjects and the average group weights of the 40 subjects categorized by their positions in the academic hierarchy.

CHAPTER V

RESULTS

Dimensions of College Teaching Effectiveness

The interpretative procedure noted above resulted in the tentative identification of four psychologically meaningful dimensions related to college teaching effectiveness. Table 3 shows the items in each dimension whose loadings met the criterion of .200 or greater. Each of the four dimensions were clearly bipolar, as the sets of items appearing at the two extremes of each dimension contrast with one another within the perspective of some identified concept of teaching effectiveness.

Dimension 1

The behaviors defining this dimension suggested the title "Showman versus Explainer." Four of the six items loading negatively represent a teacher who is good at explaining course material at a level understood by the students. The other two items, being well informed and well prepared, seemed to fit well into this pattern also. Those items loading positively do not appear to be as closely related conceptually. But when taken together, they seemed to suggest the "Showman" or the teacher concerned with making the class interesting to the student. Item 27, "Compliments

Table 3. Stimulus Behaviors and Loadings Related to the Four INDSCAL Dimensions

Dim	Item	Behavioral Item	Loading
1	26	Disregards the lowest test score of each student	.354
	12	Makes dramatic gestures and comments to emphasize important points	.302
	*27	Compliments a student on a good response	.226
	7	Explains class material clearly	-.324
	8	Is well informed about the material presented	-.315
	13	Is well prepared each day	-.301
	*17	Does not pitch his presentations above students' heads	-.218
	*21	Gives presentations that are logically arranged	-.207
	*24	Expresses concepts at a level understood by students	-.204
2	20	Encourages students to think for themselves	.305
	14	Tries to stimulate creative abilities	.293
	8	Is well informed about the material presented	.266
	26	Disregards the lowest test score of each student	-.332
	* 2	Explains how much each test counts towards the final grade	-.267
	*16	Clearly describes grading procedures	-.257
	6	Gives examples of quiz items or what to expect in a quiz	-.249
	12	Makes dramatic gestures and comments to emphasize important points	-.244
	*28	Follows course syllabus or lecture outline on schedule	-.243
	10	Uses humor that stimulates class interest and attendance	-.200
3	8	Is well informed about the material presented	.360
	*25	Clearly states the purposes and objectives of the course	.351
	20	Encourages students to think for themselves	.233
	13	Is well prepared each day	.219
	14	Tries to stimulate creative abilities	.215
	26	Disregards the lowest test score of each student	-.351
	12	Makes dramatic gestures and comments to emphasize important points	-.334
	10	Uses humor that stimulates class interest and attendance	-.282
	9	Knows or attempts to learn each student by name	-.216
4	*19	Gives tests whose content is representative of assigned material	.338
	*11	Clearly indicates what materials the tests will cover	.280
	6	Gives examples of quiz items or what to expect in a quiz	.248
	7	Explains class material clearly	.205
	9	Knows or attempts to learn each student by name	-.314
	12	Makes dramatic gestures and comments to emphasize important points	-.283
	* 1	Supplements course and text using outside references and material	-.254
	* 4	Presents problems as a challenge to the class	-.231
	14	Tries to stimulate creative abilities	-.207
	* 5	Encourages students to ask questions	-.205
	*30	Is courteous and considerate of students	-.203

*Denotes unique items to the respective dimensions.

a student on a good response," loaded uniquely on this dimension, and seems to reflect this property also.

Dimension 2

This dimension was titled "Concern for Creativity versus Concern for Structure." While the behaviors loading positively all reflect a creative orientation, four of seven of those loading negatively deal with testing. A closer look, however, suggested something other than evaluation. Items 2, 16, and 28, explaining how much each test counts, describing grading procedures, and following the course syllabus on schedule, respectively, all loaded uniquely on this dimension. These items, together with the other four, appear to have the common attribute of structure or concern for procedural detail. Thus concern for structure appeared a more appropriate descripter than evaluation.

Dimension 3

The title suggesting itself for this dimension was "Goal-directed Emphasis versus Showman." Two of the items loading negatively are the same as those loading positively on Dimension 1, and indicate more of a concern for making the class interesting than a concern for inspiring learning. The other two items, uses humor and attempts to learn students' names, tend also to fit this pattern. Thus it appeared the title "Showman" for these behaviors might be appropriate here, too. Of the items loading positively, only Item 25, "Clearly states the purposes and objectives of the course,"

is unique to this dimension. The other behaviors loading positive are compatible with this characteristic, and together suggested a goal-directed orientation towards teaching.

Dimension 4

The title selected for this dimension was "Tester versus Challenger." Whereas the majority of items loading positively represent a concern for fair and comprehensive evaluations, four of seven of the items loading negatively are unique to this dimension, and represent an orientation towards expanding the learning process beyond the course itself. Of these latter items, both presenting problems as a challenge and encouraging students to ask questions further reflect an emphasis on student involvement in rather than a lecture-type approach to classroom instruction. Thus, "Challenger" appeared appropriate.

Evaluation of the Research Hypothesis

With these tentative interpretations available as an aid, the writer proceeded to examine the clustering of subjects.

Subjects in the Stimulus Space

The 4-D INDSCAL solution produced a 40 x 4 subjects by subject weights matrix. These data are given in Table 4, along with subject numbers and corresponding plotting codes. The subject numbers correspond to organizational position as

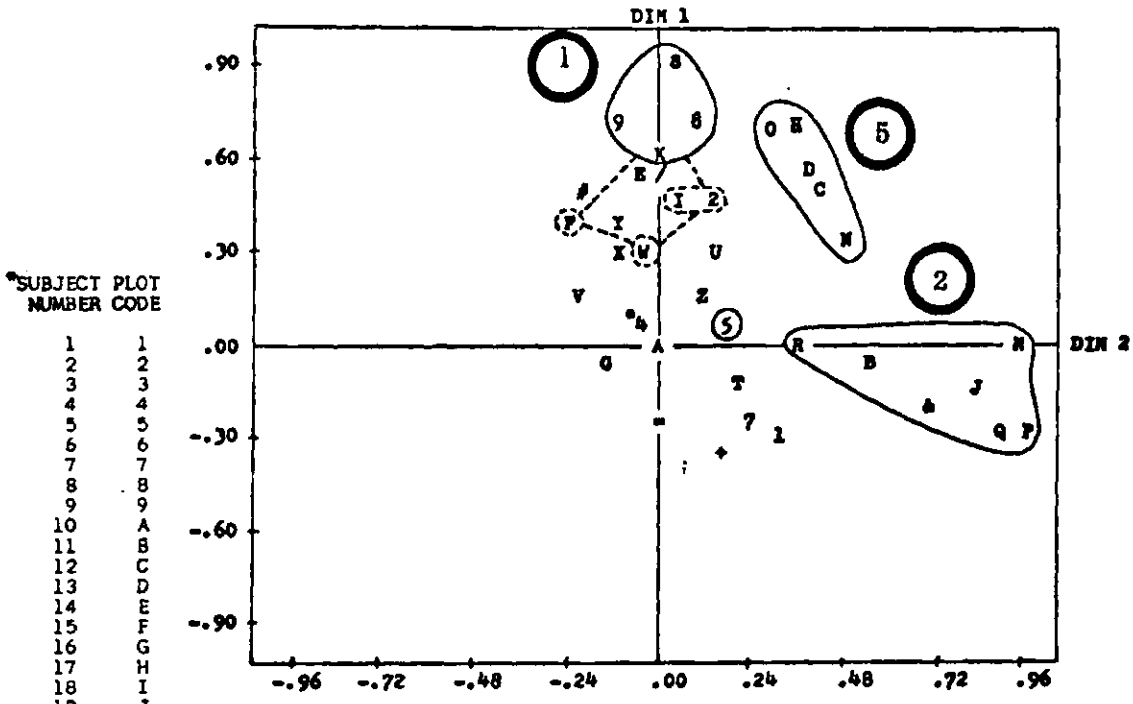
Table 4. The Subjective "Weights" Derived for Each Subject for the Four Dimensional INDSCAL Solution

Subject Number	Plot Code	Dimension 1	2	3	4
1	1	-.2833	.3223	-.0300	.6807
2	2	.4400	.1393	-.1367	-.0574
3	3	.5041	-.2052	-.1535	.3956
4	4	.0441	-.0177	-.0393	.6636
5	5	.0423	.1888	-.0712	.0456
6	6	.4951	-.2230	-.0446	.5762
7	7	-.2559	.2368	.0254	.6495
8	8	.7231	.0848	-.1430	.0565
9	9	.7179	-.1277	-.1925	.0135
10	A	.0249	.0031	-.0425	.8386
11	B	-.0700	.5449	.1598	-.0086
12	C	.4975	.4069	-.0210	-.0837
13	D	.5343	.3911	-.1227	-.0667
14	E	.5385	-.0359	.2940	-.0169
15	F	.4215	-.2478	.1185	.0730
16	G	-.0270	-.1430	.8352	.0483
17	H	.7160	.3438	-.1674	-.0758
18	I	.4299	.0379	-.0760	-.0372
19	J	-.1732	.8254	-.0739	.0793
20	K	.6064	.0051	.1727	.0482
21	L	.5056	-.2244	.0333	.2402
22	M	-.0060	.9107	-.1361	.0139
23	N	.3052	.4863	-.1370	-.0508
24	O	.6970	.2629	.0118	-.0739
25	P	-.3146	.9853	.0222	.0872
26	Q	-.2905	.8850	.0667	.1362
27	R	.0044	.3593	.1232	-.0099
28	S	.9218	.0198	-.1226	-.0045
29	T	-.1556	.2221	.2587	.0069
30	U	.3129	.1467	.2229	-.0554
31	V	.1704	-.2199	.6057	.0014
32	W	.3133	-.0328	.0959	.0778
33	X	.3167	-.0713	.4221	.0365
34	Y	.4169	-.1213	.3242	.0917
35	Z	.1610	.1071	.5028	-.0593
36	+	-.3296	.1618	.8103	.0567
37	/	.5705	.0071	.2826	-.0321
38	=	-.2516	.0037	.7326	.0615
39	*	.1240	-.0746	.7411	-.0031
40	&	-.1811	.6957	.4130	.0026

follows: 1-10 are undergraduates, 11-20 are graduates, 21-30 are teachers, and 31-40 are administrators. All 40 subjects are plotted on the basis of their subjective weights, in each of the Figures 1 through 4.

The most notable characteristic of these plots, in general, is the wide scatter of the subjects. Since the subjective weight indices, used as coordinates, reflect the relative importance attributed by each individual to the corresponding dimensions, it is clear there are marked individual differences in this respect. In Figure 1, for example, whereas subject S is located very high on Dimension 1 alone, subject M is located very high on Dimension 2 alone. Subject A falls at the origin, and subjects O, H, D, C, and N are located midway in the positive quadrant. The four plots in general clearly show that the subjects were not attributing the same importance to the four dimensions.

Two other characteristics of these plots should be noted: (1) the presence of some negative weights, and (2) a general U-shaped distribution of subjects about the axes of the positive quadrants. Both of these are likely to be effects due to the departure in the data collection procedure used here from that normally used with INDSCAL analysis, as previously noted. These effects reflect a certain degree of unreliability associated with single judgments or measures of the stimuli to the ostensibly unidimensional continuum of importance. Multiple measures of the behaviors (e.g.,



*SUBJECT PLOT NUMBER CODE

NUMBER	CODE
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	A
11	B
12	C
13	D
14	E
15	F
16	G
17	H
18	I
19	J
20	K
21	L
22	M
23	N
24	O
25	P
26	Q
27	R
28	S
29	T
30	U
31	V
32	W
33	X
34	Y
35	Z
36	+
37	/
38	=
39	*
40	&

# indicates multiple points

Figure 1. Subjects Plotted Against Dimensions 1 and 2.*

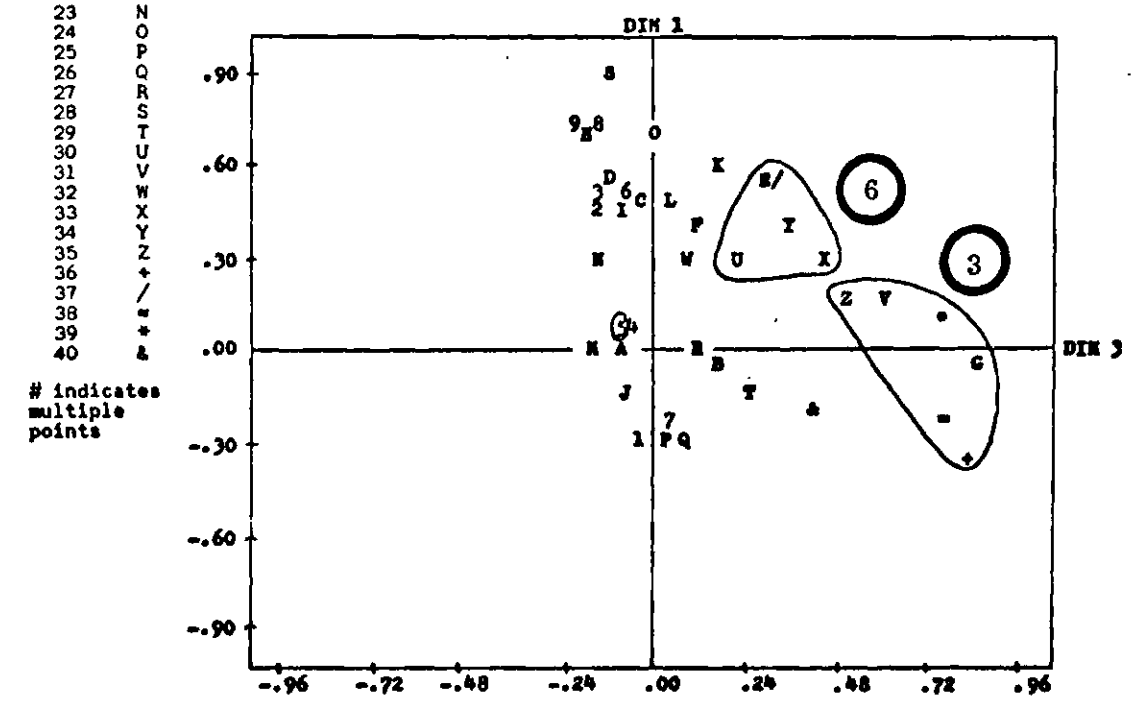


Figure 2. Subjects Plotted Against Dimensions 1 and 3.*

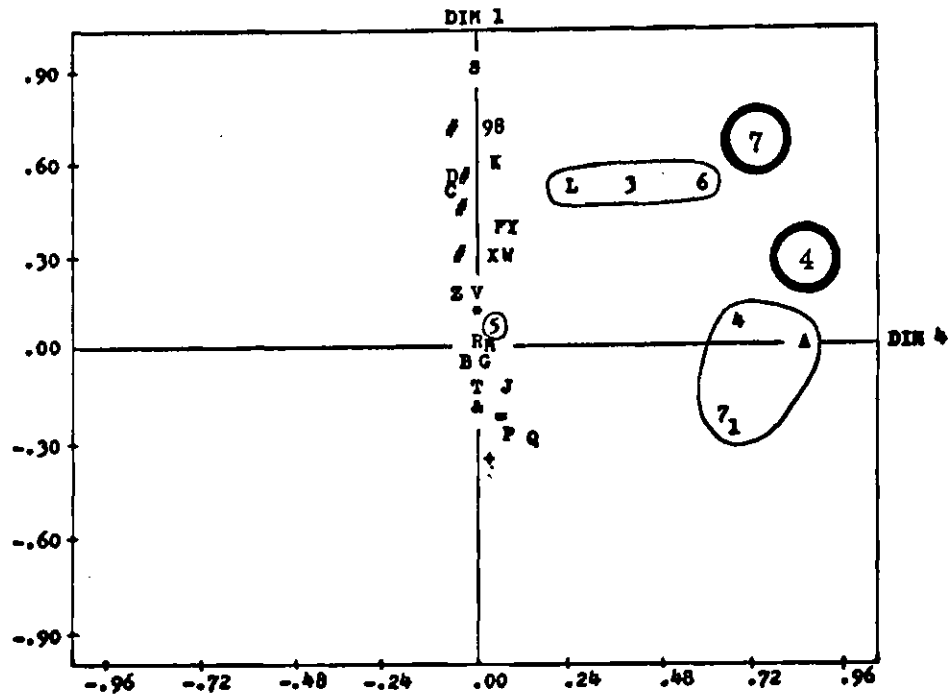


Figure 3. Subjects Plotted Against Dimensions 1 and 4.*

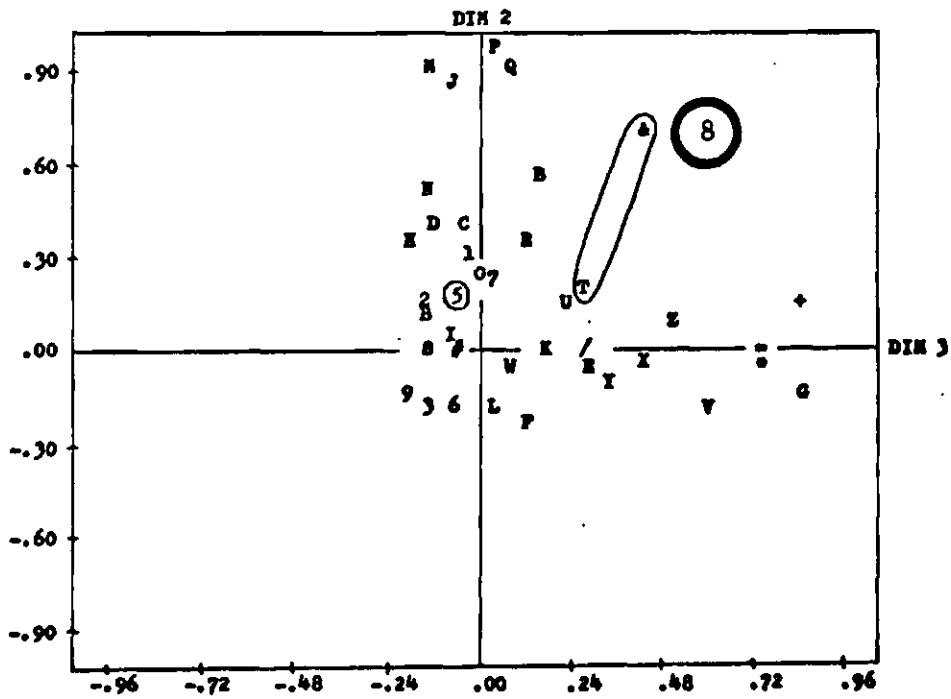


Figure 4. Subjects Plotted Against Dimensions 2 and 3.*

*SUBJECT NUMBER	PLOT CODE
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	A
11	B
12	C
13	D
14	E
15	F
16	G
17	H
18	I
19	J
20	K
21	L
22	M
23	N
24	O
25	P
26	Q
27	R
28	S
29	T
30	U
31	V
32	W
33	X
34	Y
35	Z
36	+
37	/
38	=
39	*
40	&

# indicates multiple points

paired-comparisons) would probably have resulted in a general "shift" of the subject locations away from the axes of the stimulus space and into the positive quadrants. However, as will become more evident, even with a certain degree of unreliability of measurement, clear and meaningful relationships still emerged from this research.

By examining the plots, the writer identified eight clusters of subjects. These are circled in the charts and numbered 1 to 8. It was not entirely clear from the plots where subjects 2, F, I, and W (see Figure 1) should be placed. On the basis of their weights (Table 3), they were placed in Group 1. Since subject 5 did not fit any group, he was treated by himself. While the clusters mostly suggested themselves in the plots, an analysis of the INDSCAL weights of the subjects in the respective clusters, as described below, served to justify on the basis of within group similarity and between group differences, the acceptability of these clusters.

Analysis of Subjective Weights Within Clusters

For each of the eight groups identified in the plots, the INDSCAL weights for each dimension were averaged. These values are given in Table 5. These data show that each group is readily differentiated from the others on the basis of its pattern of average weights associated with each of the four teaching dimensions. Groups 1 to 4 used only one dimension, Dimensions 1 to 4, respectively, as the sole basis

Table 5. Average "Weights" on Each of the Four INDSCAL Dimensions for Each of the Eight Groups

Group Number	Members[+]	Dimension 1	2	3	4
1	02,08,09,15, 18,20,28,32	.572[*]	-.015	-.034	.021
2	11,19,22,25, 26,27	-.142	.752[*]	.027	.050
3	16,31,35,36, 38,39	-.025	-.027	.705[*]	.018
4	01,04,07,10	-.118	.136	-.022	.708[*]
5	12,13,17,23, 24	.550[*]	.378[*]	-.087	-.070
6	14,30,33,34, 37	.431[*]	-.015	.309[*]	.005
7	03,06,21	.502[*]	-.218	-.055	.404[*]
8	29,40	-.168	.459[*]	.336[*]	.005
-	05	.042	.189	-.071	.046

[+]Members are represented by subject numbers.

[*]Dimensions attributed most value by each group.

for judging the behaviors in respect to their importance in reflecting college teaching effectiveness. Groups 5 to 8 each used various combinations of two dimensions, excluding the other two dimensions in each respective case. Subject 5 did not use any of the dimensions. He either judged the behaviors in a random fashion or in a manner related to a dimension not present in this analysis. As reflected by his low correlation of .196 (Table 2), the 4-D INDSCAL solution did not produce an adequate fit to his data. Similarly, Group 1 has a relatively lower average weight on Dimension 1 than do Groups 2, 3, and 4 on their respective dimensions of importance. This observation derives from the fact that subjects 2, 15, 18, and 32 (2, F, I, and W in Figure 1) all had only mid-range weights on Dimension 1, along with virtually zero weights on the other three dimensions. Like subject 5, they each had quite low correlations between their actual and their derived locations (ranging from .328 to .428) compared to the average subject correlation of .659 for the 4-D solution. Unlike subject 5, subjects 2, 15, 18, and 32 each possessed a pattern of weights corresponding in direction, at least, to those of the other subjects in Group 1, and thus were placed in that group.

An important feature of this group analysis is that averaging the subjective weights within clusters did not have the typical affect of masking distinct individual differences of the group members. For any given group,

inspection of the respective individual member weights (see Table 4) reveals that the average group weights per dimension are closely representative of each member's actual weights. That these clusters of individuals are distinguishable solely on the basis of the differential importance each attributed to the four perceived dimensions of college teaching effectiveness clearly supports the hypothesis of this research.

Relationships Between INDSCAL Dimensions, Groups and Stimuli

To verify the differences among groups in respect to the importance they accorded the four dimensions, the average ranking given each item in the dimensions each group weighted most heavily were computed and examined in the following manner. First the average raw score of each of the 30 behaviors was computed for each of the eight groups. Then a listing was made for each dimension of those items with loadings meeting the criterion of .200 or greater. Mean item scores were then selected for each group according to the respective dimension/s which were important for that group. These data are presented in Table 6. Subject 5 was not included for reasons previously discussed.

This analysis essentially served as a check of internal reliability of the data since one would expect there to be a direct relation between the metric solutions from INDSCAL and the actual judgments of importance made by the subjects. That is, if certain stimulus behaviors were related so as to define a dimension, and if a group of

Table 6. Mean Scores for Items Loading Highly on Each of the Four INDSCAL Dimensions, Given for Each Group Attributing Importance to the Respective Dimensions

DIM	ITEM*	GROUP							
		1	2	3	4	5	6	7	8
1	26	2.38				1.80	2.00	3.33	
	12	6.25				5.60	3.00	2.00	
	27	7.00				9.20	4.00	10.67	
	7	27.38				27.00	26.60	27.67	
	8	27.63				27.40	29.40	25.33	
	13	27.25				25.40	27.20	25.00	
	17	23.25				23.00	23.20	20.33	
	21	20.75				22.20	23.80	25.33	
	24	22.63				23.40	19.60	24.67	
2	20		28.00			26.40			25.50
	14		28.83			23.00			25.50
	8		24.33			27.40			26.00
	26		3.00			1.80			2.50
	2		5.50			3.60			8.50
	16		6.00			7.20			4.00
	6		7.33			4.60			12.50
	12		5.67			5.60			2.50
	28		6.83			5.60			5.50
	10		6.50			7.00			11.00
3	8			29.67			29.40		26.00
	25			28.83			24.60		29.50
	20			24.17			24.00		25.50
	13			23.33			27.20		20.50
	14			23.00			23.20		25.50
	26			1.83			2.00		2.50
	12			2.83			3.00		2.50
	10			3.83			7.00		11.00
	9			7.50			6.60		12.00
4	19				29.50			25.00	
	11				27.50			22.00	
	6				27.50			16.33	
	7				21.75			27.67	
	9				1.50			8.00	
	12				6.25			2.00	
	1				6.25			7.33	
	4				7.25			5.00	
	14				8.00			8.33	
	5				4.75			12.00	
	30				6.75			10.00	

* For each dimension, items loading positive are given first, then those loading negative. For each polarity, items are presented in descending order of magnitude.

subjects attributed importance to that dimension, then the magnitude of the item mean scores for that group should correspond in magnitude to the items defining the anchor points of the dimension. Those items defining one end should have very high scores and those defining the other end should have very low scores for the subjects attributing importance to the dimension.

As clearly indicated in Table 6, these expected relationships occurred without exception for each of the eight groups of subjects. This finding offers considerable credence to the validity of the relationships previously determined here.

INDSCAL Groups Versus Organizational Groups

Because subjects in rating research are normally classified or treated as homogeneous groups on the basis of organizational position, the effect of grouping subjects according to their valuing systems can be evaluated by comparing results so obtained with those obtainable when the subjects are classified according to the organizational groups they were selected to represent, i.e., undergraduates, graduates, teachers, and administrators. Table 7 contains the average weights for these initial groups on each of the four INDSCAL dimensions.

As can be seen, the relative magnitudes of the average weights for the initial groups are considerably lower than those for the INDSCAL groups, especially Groups 1 to 4 (see

Table 7. Average "Weights" on Each of the Four INDSCAL Dimensions for each of Four Intact Groups Within a Typical Academic Hierarchy

Group	Dimension 1	2	3	4
Undergraduates	.2452	.0401	-.0818	.3862
Graduates	.3474	.2128	.1119	-.0040
Teachers	.1980	.4054	.0343	.0290
Administrators	.1311	.0456	.4930	.0234
Grand Mean	.2304	.1760	.1394	.1087

Table 5). A general comparison between the two different sets of group weights illustrates, above all, how classifying subjects on the basis of their academic positions produces a marked reduction in average weights. Further, whereas the INDSCAL group weights retained the characteristic pattern of the performance perspectives held by the respective group members, as previously noted, the intact group weights masked and in some cases totally distorted the individual member perspectives. A clear example is subject 40 who is a College Dean. As reflected by the average weight of .0456 in Table 7, administrators do not seem to attribute much importance to Dimension 2 (creativity versus structure). Yet subject 40 has a weight index of .6957 on this dimension. Thus the average weight clearly does not represent the individual.

The results also illustrate the two major effects resulting from the classification of subjects according to organizational position. First, organizational classification masks for most individuals their unique perspectives, and assigns to them a composite pattern of weights that does not represent one "true", hence meaningful, perspective of the several members. For example, undergraduates had relatively high weights on Dimensions 1 and 4, graduates on 1 and 2, teachers on 2 alone, and administrators on 3 alone. Upon noting these differences, one immediately questions their meaning as being representative of the respective group

members' true value judgments in respect to the importance of the different teaching dimensions. In the light of the individual differences data analysis, it is doubtful that the differences among organizational groups denote any psychological meaning at all. For, eight distinct INDSCAL clusters were required to reflect the differing performance perspectives of the same 40 subjects. Further, with the exception of Group 4 (four undergraduates), all of the other INDSCAL groups cut across the organizational variable, as they were comprised of persons from at least two or more such groups. While the organizational group differences noted suggest a possible reason for the lack of inter-rater agreement noted in previous research, their questionable meaningfulness suggests knowledge of them per se would not be helpful in irradicating the disparities of ratings.

The second effect resulting from the classification of subjects by their organizational positions was an obvious increase in the within-group variance, and thus a consequential reduction in the between-group variance. This effect is evident from a simple examination of the range of subject weights for any given organizational group compared to the comparable range of weights for any INDSCAL group. For example, graduate students had the following ranges associated with the four respective dimensions: (1) -.1732 to .7160, (2) -.2478 to .8254, (3) -.1674 to .8352, and (4) -.0837 to .0837. INDSCAL Group 3, however, had the following

respective ranges: (1) -.3296 to .1704, (2) -.2199 to .1618, (3) .5028 to .8352, and (4) -.0554 to .0615. Whereas the INDSCAL ranges are relatively restricted on each dimension, the graduate student ranges are extensive on three of four dimensions. Since this example is quite characteristic of the two sets of groups here, it is clear that the within-group or error variance for the intact groups is exceedingly high compared to that for the INDSCAL groups. The major effect of this is, of course, to sharply reduce the between-groups variance for the organizational groups, but to increase it for the INDSCAL groups. Considering the extensive use of analysis of variance techniques in psychological research, the implications of this finding are obvious.

CHAPTER VI

DISCUSSION

The most significant finding of this thesis research was the identification of eight distinct groups of subjects on the basis of the relative importance each attributed to the behaviors comprising the four perceived dimensions of college teaching effectiveness. Four groups used only one of the dimensions, virtually excluding the other three. Four other groups used various combinations of two dimensions only. Although these groups were generally small, ranging from two to eight subjects, each was nevertheless distinct in the manner in which it attributed importance to the various teaching dimensions. Since only 40 subjects comprised the sample, it is of considerable interest that eight such distinct groups could be identified. The hypothesis that individuals from different levels within the same academic organization do not attribute importance to dimensions of college teaching effectiveness in the same manner was clearly supported.

These results have value in that they illustrate the extensive nature of individual differences in the perspectives people hold towards teaching effectiveness. They also suggest a way of characterizing subjects in terms of

important characteristics that need to be considered in assigning them to experimental and control groups used in various types of investigations relating to teacher effectiveness. For example, if in examining subject characteristics one uses the data for the organizational groups, one would say "Teachers attribute most importance to Dimension 1 whereas administrators attribute most importance to Dimension 3." But using the INDSCAL group data, one would have to say "Some administrators, some teachers, some graduate students and some undergraduate students attribute exclusive importance to Dimension 1 whereas some graduate students and some teachers attribute exclusive importance to Dimension 2." Both interpretations are correct with respect to the group data upon which they are based, both use the identical INDSCAL weights for their data base, and yet each means something completely different. The fact that such differential interpretations are possible using an identical data base suggests the need for researchers to exercise great caution in making quick and easy assumptions about the homogeneity of their sample populations.

Since the dimensions identified in this study differ in format from those usually identified in factor analyses of teaching behavior, conclusions cannot be definitive. Nevertheless the present study contributes some interesting findings concerning the restrictive use the subjects made of the several dimensions made available to them for evaluating

teaching effectiveness. Over half of the subjects attributed importance to only one dimension, while the others (except subject 5) attributed importance to various combinations of only two dimensions. While some might feel these results are a function of the single rank order given by each of the subjects, there are several factors to suggest this is not the case.

First, the 30 behaviors used here were judged to be representative of five commonly identified factors of college teaching effectiveness, so that the behavioral domain being evaluated was potentially multidimensional to each subject. Second, while the subjects were asked to rank the behaviors in reference to an ostensibly single dimension, their behavior while performing the ranking task seemed to suggest an awareness of multidimensionality. That is, in observing many of the subjects at the task, it was noted that they frequently changed their minds about which behaviors went into what categories. Further, before rank ordering the items they had initially placed in each of the five categories of importance, they shuffled the items back and forth among the categories. One professor reported he had extreme difficulty doing the task, as some of the behaviors that "did not go together as far as teaching is concerned" were still extremely important to him. For example, he said he felt it was most important for a teacher to clearly state course objectives, but it was equally important to give representative

tests. Thus, he ranked both behaviors very high, even though each was related to a different teaching dimension.

Although it cannot be determined for certain, it appears from these types of observations that the judgments made in reference to the continuum of importance represent "compactions" of the multidimensionality of the domain of teacher effectiveness. It is therefore reasonable that the four INDSCAL dimensions are actually representative of this multidimensionality, and that the subjects' locations are representative estimates of their "true" locations in respect to these same dimensions.

Regarding the dimensions themselves, it is important to emphasize that each dimension represents a relatively distinct cluster of teaching behaviors. It is the behaviors that provide the meaning, not the titles. Further, one should note that the behaviors were assessed in terms of perceived importance to effectiveness, and not in terms of actual occurrence in the classroom. Thus, the dimensions are best viewed as underlying a perceptual-evaluative space of the sample population rather than an actual teaching effectiveness space.

One finding not reported thus far concerns something that did not happen in this study. The factor of "Rapport," which has been commonly reported in factor analytic studies, did not appear in the analysis, even though five related behaviors were included here. Of these, Items 3 and 15 (see

Appendix C) did not load on any dimensions, Item 9 loaded on Dimensions 3 and 4, Item 27 on Dimension 1, and Item 30 on Dimension 4. A look at the raw scores of these items (see Appendix B) indicates they were ranked anywhere from extremely important to unimportant. However, this variable pattern is consistent with the rankings given most of the 30 behaviors by the different subjects, and thus does not suggest that differential evaluation per se was a reason for the Rapport factor not appearing.

While it is not clear exactly why this factor did not appear, one possibility is that Rapport behaviors may be perceived as coexistent with certain other performance behaviors. This could explain why the three items that did load on dimensions here did so differentially. Another possibility is that while rapport behaviors may make the learning experience more interesting and pleasureable, they may not be perceived as crucial to effective teaching. Because the rank order task prevented assigning equal importance to any of the behaviors, it is possible that rapport behaviors tended to be downgraded in lieu of other behaviors perceived as more directly related to effective teaching. There is already evidence that student evaluaters can separate the rapport factor from other effective teaching dimensions (Deshpande, et al., 1970; Solomon, et al., 1964), and that in at least one study, also performed at Georgia Tech, it was not significantly associated with effective

teaching (Deshpande, et al., 1970). But since it has been said to be associated with effective teaching and since there is a widespread belief that the "nice guy gets the good evaluations" (McKeachie, 1969), further study concerning its relevance is needed.

While the results reported here seem clear, there are at least three limitations regarding generality of the findings. First, subjects were selected so as to be representative of different academic levels rather than representative of the Institute as a whole. Second, the sample was relatively small and restricted to only one university. Third, only a single rank order of the stimulus items was obtained from each subject. Hence, before generalizing beyond the scope of this study, further replications are needed. Knowledge gained here indicates that the design of subsequent studies could be improved or sharpened by using fewer stimuli and more subjects, and by using paired-comparison data collection techniques.

Despite these limitations, the validity of the findings was strongly supported by the evidence linking the identified clusters of subjects to their patterns of judgments of the 30 stimulus behaviors. This connection between the INDSCAL solutions and the actual judgmental data clearly shows that the groups of subjects identifed reflect the perceptual-evaluative behaviors of the subjects, and are not satistical artifacts. This evidence also supports the

utility of the INDSCAL model itself in psychological research.

Because of the design of the study and the task assigned the subjects, this writer recognizes it is not proper to generalize the results to the direct rating of performance. However, consider for a moment a situation in which the subjects actually observed and evaluated the same college teacher. If ratings given should in any way be influenced by the rater's value judgments concerning the importance of various teaching behaviors, the findings of this study would lead one to expect there would be little basis for obtaining any substantial inter-rater agreement among all 40 subjects. However, considering the favorable reduction of within-group variance for the INDSCAL groups, one could feasibly hypothesize that inter-rater agreement between subjects sharing common perspectives would be high compared to that for subjects holding different and/or opposing perspectives towards the same dimensions. Moreover, if teachers could be effectively "scored" on a set of dimensions, one could investigate teacher by rater interactions related to the teacher evaluation process. For example, one could predict that a creatively oriented teacher would receive very low ratings from a rater who attributes considerable importance to the evaluative function of teaching. Similarly, one could investigate the extent to which teacher effectiveness is evaluated as a function of what a teacher actually does versus what an observer <u>thinks</u> he

should be doing.

In view of the finding that over half of the subjects attributed importance to only one dimension, and that all but one of the remaining subjects attributed importance to various combinations of only two dimensions, the answer to Ronan & Schwartz's (1974) question, as to just what it is that raters are rating, appears to be "a very limited dimensionality of performance." For the finding suggests that persons observing any given job performance attend to and remember those few critical behaviors that relate directly to job dimensions which to them are valued as important. If this proved to be the correct answer, then it would appear that one difficulty with ratings is that the different raters may have focused on different and restricted performance dimensions.

In a more general way, the results of this study raise a question concerning why or how individuals come to value certain dimensions to the exclusion of others. The findings further suggest the possibility of identifying some physchological bases underlying the manner in which certain people come to commonly value specific dimensions of performance over others. Wish (1971), for example, has suggested from his INDSCAL research that individual conceptions towards nations have as their basis the "hawk" versus "dove" attitudes people generally hold regarding political or power alignments of governments. Whether or not attitudes or other

such personal characteristics underlie individual performance perspectives remains to be seen. Since we are now beginning to document the existence of basic individual differences such as those found here, however, it appears some of the answers to the questions of causation of performance perspectives are potentially close at hand.

Finally, the findings of this study offer general but indirect support for Guion's (1965) contentions regarding the causal bases underlying the disparities of ratings. More specifically, they offer some support for his contention that raters from different organizational levels attribute differential importance to dimensions of job performance. Although the subjects here did not rate actual performance, they did indicate what to them were the most important behaviors related to the domain of college teaching effectiveness. From the analysis of their value judgments, it was shown that different clusters of individuals were characterized by their differential ascriptions of importance to different dimensions of that domain. But, the basis for common performance perspectives was not determined by "position" in the academic institute. Categorizing subjects by their organizational positions markedly reduced the magnitude of individual differences that existed between the eight different INDSCAL clusters of subjects.

Even though the present study was not intended to investigate the causal bases underlying disparities of

ratings, it has suggested several questions that merit investigation. Hopefully research stimulated by the results of this study will assist in identifying the causal bases underlying the disparities of performance ratings.

APPENDICES

APPENDIX A

THE EXPERIMENTAL TASK

Part 1*

Instructions

Please provide appropriate information for each of the following:

Age: _____ Sex: Male _____

Height: _____ Female _____

Weight: _____

Position or Title: _____

Professional Field: _____

Years of Formal Education: _____

Following are five general areas of responsibility of college classroom teachers. Please rank order of these regarding their relative importance in determining a teacher's classroom effectiveness. Use the numbers 1 through 5: 1 = the most important area, 2 = the next most important area, etc., through 5 = the least important area.

_____ Clearness of grading objectives and procedures.

_____ Organization and structure of course content; mastery of subject and course content.

_____ Rapport with students; helpful attitude.

_____ Correct estimation of student abilities; assignment of work commensurate with student abilities.

_____ Stimulation and inspiration of creative and critical thinking.

---

*This was the form used for teachers and administrators. That used for graduates and undergraduates differed only in that it contained a checklist for "Year in School" instead of "Position or Title."

Part 2*

Instructions

1. The following two pages contain 30 teacher behaviors. Lay the pages in front of the box, and read each statement. Think how you would feel if you actually saw a college teacher performing that behavior, and decide the degree of importance it would have regarding teacher effectiveness.

2. After reading all of the statements, remove each one from its sheet and place it in one of the five categories in the box.

   a. A means little or no importance.
   b. B means some importance.
   c. C means quite a bit of importance.
   d. D means an extreme amount of importance.
   e. E means an extraordinary amount of importance.

3. In making your judgements, place six statements in each category. For your convenience, the last page of this part contains a Tally Sheet so you can keep account of the number of statements you place in each category. You may change the statements around as much as you want, but when you finish you must have six in each category.

4. After you have placed six statements in each category, remove those from category E. Then, rank them in descending order of importance from "most to least," and number them from 1 to 6 (1 = most, to 6 = least important). Paper clip them together and return them to their box compartment.

5. Repeat step 4 for each of the remaining categories D, C, B, and A. After you have finished ranking the six statements within each category, you may go on to Part III.

---

*Following this instruction sheet were two pages containing the 30 behavioral statements to be rank ordered.

Part 3*

Instructions

This is a questionnaire to find out the way in which certain important events in our society affect different people. Each item consists of a pair of alternatives lettered a and b. Please circle the one statement of each pair (and only one) which you more strongly believe to be the case as far as you are concerned. Be sure to circle the letter of the one you actually believe to be more true rather than the one you think you should choose or the one you would like to be true ideally. This is a measure of your personal belief, obviously there are no right or wrong answers.

Please answer these items carefully, but do not spend too much time on any one item. In some cases you may discover that you believe both statements or neither statement. However, in all cases select the one statement of each pair that you more strongly believe to be the case as far as you are concerned.

1. a. Children get into trouble because their parents punish them too much.
   b. The trouble with most children nowadays is that their parents are too easy with them.
2. a. Many of the unhappy things in people's lives are partly due to bad luck.
   b. People's misfortunes result from the mistakes they make.
3. a. One of the major reasons why we have wars is because people don't take enough interest in politics.
   b. There will always be wars, no matter how hard people try to prevent them.
4. a. In the long run people get the respect they deserve in this world.
   b. Unfortunately, an individual's worth often passes unrecognized no matter how hard he tries.
5. a. The idea that teachers are unfair to students is nonsense.
   b. Most students don't realize the extent to which their grades are influenced by accidental happenings.
6. a. Without the right breaks one cannot be an effective leader.
   b. Capable people who fail to become leaders have not taken advantage of their opportunities.
7. a. In the case of the well prepared student there is rarely if ever such a thing as an unfair test.
   b. Many times exam questions tend to be so unrelated to course work that studying is really useless.

*Shortened form of the Internal-External Locus of Control Scale.

8. a. Becoming a success is a matter of hard work, luck has little or nothing to do with it.
   b. Getting a good job depends mainly on being in the right place at the right time.

9. a. The average citizen can have an influence in government decisions.
   b. This world is run by the few people in power, and there is not much the little guy can do about it.

10. a. When I make plans, I am almost certain that I can make them work.
    b. It is not always wise to plan too far ahead because many things turn out to be a matter of good or bad luck anyway.

11. a. In my case getting what I want has little or nothing to do with good luck.
    b. Many times we might just as well decide what to do by flipping a coin.

12. a. Who gets to be the boss often depends on who was lucky enough to be in the right place first.
    b. Getting people to do the right thing depends upon ability, luck has little or nothing to do with it.

13. a. As far as world affairs are concerned, most of us are the victims of forces we can neither understand nor control.
    b. By taking an active part in political and social affairs, the people can control world events.

14. a. Most people don't realize the extent to which their lives are controlled by accidental happenings.
    b. There really is no such thing as "luck."

15. a. It is hard to know whether or not a person really likes you.
    b. How many friends you have depends upon how nice a person you are.

16. a. With enough effort we can wipe out political corruption.
    b. It is difficult for people to have much control over the things politicians do in office.

17. a. Sometimes I can't understand how teachers arrive at the grades they give.
    b. There is a direct connection between how hard I have tried and the grades I have gotten.

18. a. Many times I feel that I have little influence over the things that happen to me.
    b. It is impossible for me to believe that chance or luck plays any role in my life.

19. a. What happens to me is always my own doing.
    b. Sometimes I feel that I don't have enough control over the direction of my life.

APPENDIX B

RAW DATA

SUBJECT	ITEMS																													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	1	23	11	5	3	30	10	15	2	8	26	7	14	9	16	20	17	10	28	13	21	27	24	12	25	29	6	22	19	4
2	12	15	22	11	21	3	29	24	25	7	19	13	30	27	8	2	17	26	20	28	9	18	16	23	4	1	14	5	10	6
3	8	14	27	4	13	20	28	30	9	11	21	1	29	5	26	19	16	15	23	12	24	25	22	17	6	7	10	18	3	2
4	4	17	21	10	7	29	27	18	1	8	26	11	15	13	3	12	23	24	30	19	22	20	14	25	6	28	2	5	16	9
5	7	5	26	14	21	20	22	30	8	3	25	9	12	29	27	6	17	23	28	18	4	10	11	16	2	19	24	1	15	13
6	5	19	18	9	15	22	26	20	7	12	23	1	16	6	8	14	30	17	27	21	28	25	24	29	13	2	4	3	10	11
7	16	25	9	11	2	28	18	22	1	7	29	6	17	5	8	26	12	14	30	10	23	27	21	15	24	19	20	13	3	4
8	5	7	27	10	20	12	30	25	8	2	14	1	18	23	24	6	29	26	17	22	11	19	21	28	15	4	3	13	16	9
9	4	7	24	26	12	15	30	27	10	21	14	2	25	11	23	8	28	19	17	13	26	18	16	29	9	1	3	20	5	22
10	4	19	16	3	7	23	24	22	2	9	29	1	15	5	12	20	26	28	30	6	18	17	25	27	11	14	8	13	21	10
11	13	9	19	29	12	7	28	16	3	5	6	1	10	30	11	8	18	22	14	24	20	21	15	27	17	2	25	4	23	26
12	6	4	20	17	22	9	26	25	11	7	2	3	28	29	21	10	27	24	13	30	16	18	14	23	12	1	15	5	19	8
13	26	3	20	15	16	2	29	30	9	5	8	6	24	21	19	7	23	13	17	22	28	27	18	14	12	1	11	4	25	10
14	16	13	24	18	6	8	28	30	3	11	9	2	27	26	7	14	22	15	20	25	29	17	12	21	19	1	4	10	5	23
15	11	17	14	20	13	15	28	30	25	4	6	7	29	9	3	21	24	18	16	19	27	10	22	26	23	8	2	12	5	1
16	4	16	14	21	10	7	20	30	13	2	11	3	24	22	9	27	19	12	28	23	26	18	5	17	29	1	8	6	25	15
17	18	1	21	17	16	7	28	25	10	3	15	8	27	20	22	5	23	19	9	29	24	30	12	26	13	2	4	6	14	11
18	19	3	14	6	16	13	28	29	2	25	8	21	30	22	10	7	18	5	20	23	24	27	15	17	11	1	4	12	26	9
19	25	3	26	22	29	1	21	23	4	7	9	13	5	28	24	8	17	16	10	30	19	14	11	20	18	2	6	15	27	12
20	21	8	14	15	12	9	22	30	4	11	19	1	28	20	13	7	27	23	25	29	24	17	6	26	18	2	3	10	16	5
21	9	12	10	2	20	7	29	26	8	5	22	4	30	14	15	21	19	11	25	3	24	27	23	28	13	1	18	6	16	17
22	26	7	24	30	23	6	19	29	3	12	11	2	25	28	22	1	15	13	8	27	14	20	17	18	10	5	9	4	21	16
23	2	5	28	20	21	1	25	29	12	17	11	10	18	19	27	7	16	23	13	26	14	24	6	30	9	3	8	4	15	22
24	11	5	21	19	15	4	27	28	6	3	13	1	30	26	17	7	23	10	14	25	29	20	18	24	12	2	8	9	16	22
25	26	5	25	29	22	10	23	24	7	4	9	8	12	28	20	6	15	11	13	30	16	18	3	19	14	1	17	2	27	21
26	27	7	21	26	23	3	9	30	15	10	8	2	24	29	20	4	5	16	12	28	18	13	14	22	17	1	11	6	19	25
27	27	2	21	4	28	17	15	24	13	1	16	8	26	30	14	9	19	25	5	29	12	23	3	18	22	7	11	10	6	20
28	9	6	22	16	19	5	28	26	3	4	15	2	29	25	21	7	30	23	20	14	27	17	18	24	12	1	11	8	10	13
29	11	15	28	14	27	21	19	22	16	12	13	2	17	24	7	1	9	18	8	23	20	29	3	10	30	4	26	5	6	25
30	5	19	18	23	21	13	28	29	20	4	10	1	27	25	22	6	11	15	14	30	26	12	16	17	24	2	9	3	7	8
31	3	13	6	17	22	11	26	30	10	8	12	7	25	18	5	15	19	21	14	23	27	24	2	20	29	1	9	28	16	4
32	12	19	23	4	17	15	24	30	5	14	20	3	29	27	22	25	13	9	21	26	18	28	10	8	6	1	16	2	11	7
33	19	12	17	13	3	8	20	28	1	5	18	7	29	26	6	14	15	23	10	27	22	25	21	24	30	4	2	11	9	16
34	4	9	13	12	15	10	28	30	6	5	21	1	29	20	19	16	26	8	17	18	14	27	25	11	23	2	3	7	22	24
35	13	7	12	8	25	4	23	29	9	5	3	1	28	24	10	6	14	16	19	26	22	27	15	20	30	2	21	18	11	17
36	5	7	15	28	27	6	11	29	2	1	20	4	14	24	8	19	16	21	26	23	12	22	10	17	30	3	13	9	18	25
37	11	5	17	15	18	6	29	30	3	10	7	4	24	19	9	14	23	26	8	20	28	22	16	25	27	1	2	13	21	12
38	3	15	19	26	9	21	17	30	7	4	18	1	20	23	16	22	10	11	25	24	8	29	5	12	27	2	13	6	14	28
39	15	8	10	21	17	9	24	30	4	3	13	1	29	27	11	19	23	22	25	26	12	18	6	16	28	2	7	20	14	5
40	12	2	16	26	22	4	19	30	8	10	5	3	24	27	17	7	15	23	9	28	11	21	14	20	29	1	13	6	25	18

APPENDIX C

STIMULUS BEHAVIORS WITH INDSCAL COEFFICIENTS

	VARIABLE TITLES	O-1	O-2	O-3	O-4
1	SUPPLEMENTS COURSE AND TEXT USING OUTSIDE REFERENCES AND MATERIAL	.112	.103	-.171	-.254
2	EXPLAINS HOW MUCH EACH TEST COUNTS TOWARDS THE FINAL GRADE	.192	-.267	-.113	.110
3	WILLINGLY GIVES INDIVIDUAL ASSISTANCE	-.109	.181	-.043	.006
4	PRESENTS PROBLEMS AS A CHALLENGE TO THE CLASS	.036	.192	.116	-.231
5	ENCOURAGES STUDENTS TO ASK QUESTIONS	-.010	.171	.027	-.205
6	GIVES EXAMPLES OF QUIZ ITEMS OR WHAT TO EXPECT IN A QUIZ	.176	-.249	-.151	.248
7	EXPLAINS CLASS MATERIAL CLEARLY	-.324	.143	.130	.205
8	IS WELL INFORMED ABOUT THE MATERIAL PRESENTED	-.315	.266	.360	.137
9	KNOWS OR ATTEMPTS TO LEARN EACH STUDENT BY NAME	.196	-.170	-.216	-.314
10	USES HUMOR THAT STIMULATES CLASS INTEREST AND ATTENDANCE	.196	-.200	-.282	-.174
11	CLEARLY INDICATES WHAT MATERIALS THE TESTS WILL COVER	.053	-.162	-.067	.280
12	MAKES DRAMATIC GESTURES AND COMMENTS TO EMPHASIZE IMPORTANT POINTS	.302	-.244	-.354	-.283
13	IS WELL PREPARED EACH DAY	-.301	.103	.219	.044
14	TRIES TO STIMULATE CREATIVE ABILITIES	-.130	.293	.215	-.207
15	TAKES TIME TO HELP STUDENTS AFTER CLASS	-.030	.112	-.152	-.111
16	CLEARLY DESCRIBES GRADING PROCEDURES	.144	-.257	.047	.093
17	DOES NOT PITCH HIS PRESENTATIONS ABOVE STUDENTS HEADS	-.218	.019	.040	.152
18	RECOGNIZES STUDENTS LIMITATIONS IN UNDERSTANDING NEW MATERIAL	-.078	.052	.055	.088
19	GIVES TESTS WHOSE CONTENT IS REPRESENTATIVE OF ASSIGNED MATERIAL	-.041	-.105	.123	.338
20	ENCOURAGES STUDENTS TO THINK FOR THEMSELVES	-.163	.305	.233	-.080
21	GIVES PRESENTATIONS THAT ARE LOGICALLY ARRANGED	-.207	.036	.080	.163
22	STRESSES IMPORTANT POINTS AND GENERAL CONCEPTS	-.151	.088	.169	.194
23	ASSIGNS A REASONABLE AMOUNT OF WORK	-.026	-.092	-.159	.164
24	EXPRESSES CONCEPTS AT A LEVEL UNDERSTOOD BY STUDENTS	-.204	.139	.048	.155
25	CLEARLY STATES THE PURPOSES AND OBJECTIVES OF THE COURSE	.022	-.006	.351	-.026
26	DISREGARDS THE LOWEST TEST SCORE OF EACH STUDENT	.354	-.332	-.351	.033
27	COMPLIMENTS A STUDENT ON A GOOD RESPONSE	.226	-.079	-.131	-.181
28	FOLLOWS COURSE SYLLABUS OR LECTURE OUTLINE ON SCHEDULE	.154	-.243	-.095	-.094
29	WELCOMES DIFFERENT VIEWPOINTS	.056	.129	.019	-.050
30	IS COURTEOUS AND CONSIDERATE OF STUDENTS	.089	.066	.035	-.203

APPENDIX D

STIMULUS ITEM MEAN SCORES FOR EIGHT INDSCAL GROUPS

ITEM	GROUP 1 N=8	GROUP 2 N=6	GROUP 3 N=6	GROUP 4 N=4	GROUP 5 N=5	GROUP 6 N=5	GROUP 7 N=3	GROUP 8 N=2	SUBJECT 5	GRAND MEAN N=40
1	11.625	24.000	7.167	6.250	12.600	11.000	7.333	11.500	7.000	11.875
2	10.250	5.500	11.000	21.000	3.600	11.600	15.000	8.500	5.000	10.200
3	20.000	22.667	12.667	14.250	22.000	17.800	18.333	22.000	26.000	18.825
4	13.500	23.333	20.167	7.250	17.600	16.200	5.000	20.000	14.000	15.900
5	16.250	22.833	18.333	4.750	18.000	12.600	16.000	24.500	21.000	16.675
6	10.875	7.333	9.667	27.500	4.600	9.000	16.333	12.500	20.000	11.525
7	27.375	19.167	20.167	21.750	27.000	26.600	27.667	19.000	22.000	23.825
8	27.625	24.333	29.667	19.250	27.400	29.400	25.333	26.000	30.000	26.600
9	10.250	7.500	7.500	1.500	9.600	6.600	8.000	12.000	8.000	7.875
10	11.000	6.500	3.833	8.000	7.000	7.000	9.333	11.000	3.000	7.625
11	14.375	9.833	12.833	27.500	9.800	13.000	22.000	9.000	25.000	14.600
12	6.250	5.667	2.833	6.250	5.600	3.000	2.000	2.500	9.000	4.725
13	27.250	17.000	23.333	15.250	25.400	27.200	25.000	20.500	12.000	22.800
14	20.500	28.833	23.000	8.000	23.000	23.200	8.333	25.500	29.000	21.075
15	15.500	18.500	9.833	9.750	21.200	12.600	16.333	12.000	27.000	15.050
16	10.375	6.000	18.000	19.500	7.200	12.800	18.000	4.000	6.000	11.825
17	23.250	14.833	16.833	19.500	22.400	19.400	21.667	12.000	17.000	19.225
18	18.625	17.167	17.167	19.000	17.800	17.400	14.333	20.500	23.000	17.850
19	19.500	10.333	22.833	29.500	13.200	13.800	25.000	8.500	28.000	18.200
20	21.750	28.000	24.167	12.000	26.400	24.000	12.000	25.500	18.000	22.300
21	20.750	16.500	17.833	21.000	22.200	23.800	25.333	15.500	4.000	19.925
22	19.250	18.167	23.000	22.750	23.800	20.600	25.667	25.000	10.000	21.275
23	15.500	10.500	7.167	21.000	13.600	18.000	23.000	8.500	11.000	14.225
24	22.625	20.667	17.000	19.750	23.400	19.600	24.667	15.000	16.000	20.525
25	12.250	16.333	28.833	16.500	11.600	24.600	10.667	29.500	2.000	17.725
26	2.375	3.000	1.833	22.500	1.800	2.000	3.333	2.500	19.000	4.775
27	7.000	13.167	11.833	9.000	9.200	4.000	10.667	19.500	24.000	10.075
28	10.250	6.833	14.500	13.250	5.600	8.800	9.000	5.500	1.000	9.350
29	12.375	20.500	16.333	14.750	17.800	12.800	9.667	15.500	15.000	15.175
30	9.000	20.000	15.667	6.750	14.600	16.600	10.000	21.500	13.000	13.875

BIBLIOGRAPHY

Anastasi, A. Psychological testing. (3rd ed.) New York: Macmillan, 1968.

Anastasi, A. The concept of validity in the interpretation of test scores. Educational and Psychological Measurements, 1950, 10, 67-78.

Barrett, R. S. Performance rating. Chicago: Science Research Associates, 1966.

Basco, R. O. and Lawshe, C. H. Foreman leadership as perceived by superiors and subordinates. Personnel Psychology, 1959, 12, 573-582.

Beckman, L. Effects of students' performance on teachers' and observers' attribution of casuality. Journal of Educational Psychology, 1970, 61, 76-82.

Blum, D. and Naylor, J. Industrial psychology: Its theoretical and social foundations. (Rev. ed.) New York: Harper and Row, 1968.

Borg, W. R. and Hamilton, E. R. Comparison between a performance test and criteria of instructor effectiveness. Psychological Reports, 1956, 2, 111-116.

Braunstein, D. N. and Benston, G. J. Student and department chairman views of the performance of university professors. Journal of Applied Psychology, 1973, 58, 244-249.

Buckner, D. N. The predictability of ratings as a function of interrater agreement. Journal of Applied Psychology, 1959, 43, 60-64.

Callahan, C. M. and Messé, L. A. Conditions affecting attempts to convert fate control to behavior control. Journal of Experimental Research in Personality, 1973, 3, 481-490.

Campbell, J. P., Dunnette, M. D., Lawler, E. E., and Weick, K. E. Managerial behavior, performance and effectiveness. New York: McGraw-Hill, 1970.

Carroll, J. D. and Chang, J. J. Analysis of individual differences in multidimensional scaling via N-way generalization of Eckart-Young decomposition. Psychometrika, 1970, 35, 283-319.

Coffman, W. E. Determining students' concepts of effective teaching from their ratings of instructors. Journal of Educational Psychology, 1954, 45, 277-286.

Crawford, P. L. and Bradshaw, H. L. Perception of characteristics of effective univeristy teachers: A scaling analysis. Educational and Psychological Measurement, 1968, 28, 1079-1085.

Deshpande, A. S. and Webb, S. C. and Marks, E. Student perceptions of engineering instructor behaviors and their relationships to the evaluation of instructors and courses. American Educational Research Journal, 1970, 7, 289-305.

Eiser, J. R. Judgement of attitude statements as a function of judges' attitudes and the judgmental dimension. British Journal of Social and Clinical Psychology, 1973, 12, 231-240.

Friedlander, F. Performance and interactional dimensions of organizational work groups. Journal of Applied Psychology, 1966, 50, 257-265.

Ghiselli, E. E. Differentiation of individuals in terms of their predictability. Journal of Applied Psychology, 1956, 40, 374-377.

Graham, W. K. and Calendo, J. T. Personality correlates of supervisory ratings. Personnel Psychology, 1969, 22, 483-488.

Green, P. E. On the robustness of multidimensional scaling techniques. Journal of Marketing Research, 1975, 12, 73-81.

Green, P. E. and Carmone, F. J. Multidimensional scaling and related techniques. Boston: Allyn and Bacon, 1970.

Green, P. E. and Rao, V. R. Applied multidimensional scaling. New York: Holt, Rinehart and Winston, 1972.

Gruenfeld, L. W. and Weissenberg, P. Relationship between supervisory cognitive style and social orientation. Journal of Applied Psychology, 1974, 59, 386-388.

Guion, R. Personnel testing. New York: McGraw-Hill, 1965.

Isaacson, R. L., McKeachie, W. J. and Milholland, J. E. Correlation of teacher personality variables and student ratings. Journal of Educational Psychology, 1963, 4, 110-117.

Klimorski, R. J. and London, M. Role of the rater in performance appraisal. Journal of Applied Psychology, 1974, 59, 445-451.

Maslow, A. H. and Zimmerman, W. College teaching ability, scholarly activity and personality. Journal of Educational Psychology, 1956, 47, 185-189.

McClelland, J. N. and Rhodes F. Prediction of job success for hospital aides and orderlies from MMPI scores and personal history data. Journal of Applied Psychology, 1969, 54, 49-54.

McCormick, E. J. and Tiffin, J. Industrial psychology. (6th ed.) Englewood Cliffs, N. J.: Prentice-Hall, 1974.

McKeachie, W. J. Teaching tips: A guidebook for the beginning college teacher. (6th ed.) Lexington, Mass: D. C. Heath, 1969.

Odiorne, G. S. The need for an economic approach to training. Journal of American Society for Training Directors, 1964, 18, 3-12.

Prien, E. P. Assessments of high-level personnel: V. An analysis of interviewers' prediction of job performance. Personnel Psychology, 1962, 15, 319-334.

Roach, D. E. and Wherry, R. J. Performance dimensions of multiline insurance agents. Personnel Psychology, 1970, 23, 239-250.

Rodin, M. and Rodin, B. Student evaluations of teachers. Science, 1972, 177, 1164-1166.

Ronan, W. W. A factor analysis of eight job performance measures. Journal of Industrial Psychology, 1963, 1, 107-112. (a)

Ronan, W. W. A factor analysis of eleven job performance measures. Personnel Psychology, 1963, 16, 255-267. (b)

Ronan, W. W. Evaluation of three criteria of management performance. Journal of Industrial Psychology, 1970, 5, 18-20.

Ronan, W. W. Development of an instrument to evaluate college classroom teaching effectiveness. (Project No. 1-D-045, Grant No. OEG-4-71-0067) Georgia Tech: Office of Education, 1971.

Ronan, W. W. and Prien, E. P. Toward a criterion theory: A review and analysis of research and opinion. Richardson Foundation, 1966.

Ronan, W. W. and Prien, E. P. Perspectives on the measurement of human performance. New York: Appleton-Century-Crofts, 1971.

Ronan, W. W. and Latham, G. P. The reliability and validity of the critical incident technique: A closer look. Studies in Personnel Psychology, 1974, 6, 53-64.

Ronan, W. W. and Prien, E. P. Performance evaluation and criterion construction. Unpublished manuscript, Georgia Institute of Technology, 1974.

Ronan, W. W. and Schwartz, A. P. Ratings as performance criteria. International Review of Applied Psychology, 1974, 23, 71-83.

Ryans, D. G. and Frederickson, N. Performance test of educational achievement. In E. F. Lindquist (Ed.), Educational measurement. Washington, D. C.: American Council on Education, 1951.

Schmidt, F. L. and Kaplan, L. B. Composite versus multiple criteria: A review and resolution of the controversy. Personnel Psychology, 1971, 24, 419-434.

Schrauger, S. and Altrocchi, J. The personality of the perceiver as a factor in person perception. Psychological Bulletin, 1964, 62, 289-308.

Schwartz, A. P., Ronan, W. W. and Day, G. J. Individual differences and job satisfaction. Studies in Personnel Psychology, 1975, in press.

Smith, P. C. and Kendall, J. G. Retranslation of expectations: An approach to the construction of unambiguous anchoring of rating scales. Journal of Applied Psychology, 1963, 47, 149-155.

Solomon, D. Teacher behavior dimensions, course characteristics, and student evaluations of teachers. American Educational Research Journal, 1966, 3, 35.

Solomon, D., Rosenberg, L. and Bezdek, W. E. Teacher behavior and student learning. Journal of Educational Psychology, 1964, 55, 23-30.

Springer, D. Ratings of candidates for promotion for coworkers and supervisors. Journal of Applied Psychology, 1953, 37, 347-351.

Stander, N. E. A longitudinal study of some relationships among criteria of managerial performance as perceived by superiors and subordinates. Journal of Industrial Psychology, 1965, 3, 43-51.

Stogdill, R. M., Shartle, C., Wherry, R. J. and Jaynes, W. A factor analytic study of administrative behavior. Personnel Psychology, 1955, 8, 165-180.

Thompson, R. P. The characteristics of teaching behavior as measured by a student rating form. Master's Thesis, University of Georgia, Athens, 1972.

Thorndike, R. L. Personnel selection. New York: John Wiley and Sons, 1949.

Torgerson, W. S. Theory and methods of scaling. New York: John Wiley and Sons, 1958.

Turner, W. W. Dimensions of foremen performance: A factor analysis of criterion measures. Journal of Applied Psychology, 1960, 44, 316-323.

Tversky, A. and Krantz, D. H. Similarity of schematic faces: A test of interdimensional additivity. Perception and Psychophysics, 1969, 5, 124-128.

Wherry, R. J. The past and future of criterion evaluation. Personnel Psychology, 1957, 10, 1-5.

Wish, M. Individual differences in perceptions and preferences among nations. In C. W. King and D. Tigert (Eds.) Attitude research reaches new heights. Chicago: American Marketing Association, 1971.

Zedeck, S., Imparato, N., Krausz, M. and Oleno, T. Development of behaviorally anchored rating scales as a function of organizational level. Journal of Applied Psychology, 1974, 59, 249-252.